

Answer the following questions:**Question No. 1****(10 marks)**

Put (✓) in front of the right statement and (×) in front of the wrong one, then correct it.

1. Let f be some arbitrary/unknown function so that $f(\theta_0, \theta_1)$ outputs a number (f is not necessarily the cost function of linear regression, so it may have local optima). Suppose we use gradient descent to try to minimize $f(\theta_0, \theta_1)$ as a function of θ_0 and θ_1 . If θ_0 and θ_1 are initialized at a local minimum, the one iteration will change their values.
2. You are training a classification model with logistic regression. Adding a new feature to the model always results in equal or better performance on examples not in the training set.
3. A* search algorithm is a depth-first search with $h(n) = 0$.
4. Breadth-first graph search is guaranteed to return an optimal solution.
5. Consider two different A* heuristics $h_1(s)$, and $h_2(s)$ that are each admissible. Combine the two heuristics into a single heuristic, using some function g . if $g = \max(h_1(s), h_2(s))$ that will result in A* guaranteed to find the optimal solution while still guaranteeing admissibility.
6. A* graph search heuristics consistency implies admissibility.
7. Forward checking detects failure earlier than arc consistency.
8. You run gradient descent for 15 iterations with $\alpha=0.3$ and compute $J(\theta)$ after each iteration. You find that the value of $J(\theta)$ decreases slowly and is still decreasing after 15 iterations. Based on this, Rather than use the current value of α , it'd be more promising to try a larger value of α .
9. Goal driven reasoning or backward chaining is an inference technique which uses IF THEN rules to deduce a problem solution from initial data.
10. The greedy best-first search algorithm is complete.

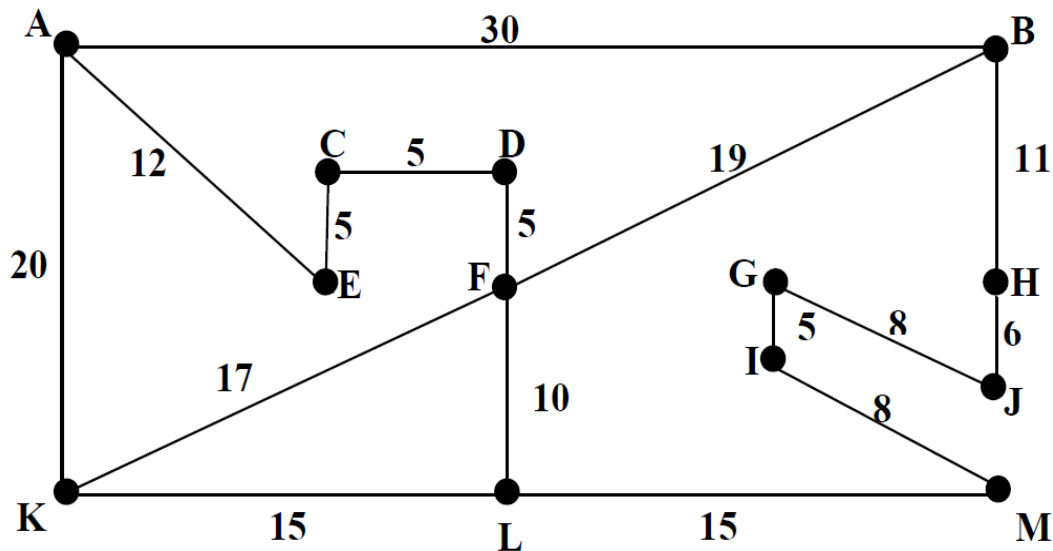
Question No. 2**(20 marks)**

1. Compare **in a table** between breadth-first, depth first and uniform-cost search algorithms (Discuss five comparison points).

2. Consider the following map (not drawn to scale).

Use the A* algorithm to work out a route from town A to town M. Use the following cost functions.

- $G(n)$ = The cost of each move as the distance between each town (shown on map).
- $H(n)$ = The Straight Line Distance between any town and town M. These distances are given in the table below.



Straight Line Distance to M

A	B	C	D	E	F	G	H	I	J	K	L	M
56	22	30	29	29	30	14	10	8	5	30	15	0

- Provide the search tree for your solution, showing the order in which the nodes were expanded and the cost at each node. You should not re-visit a town that you have just come from. State the route you would take and the cost of that route.
- Assume the estimated costs by the heuristic were replaced and shown in the following table

Straight Line Distance to M

A	B	C	D	E	F	G	H	I	J	K	L	M
80	10	50	20	10	30	60	30	20	50	60	20	0

What route would now be returned by the A* algorithm and what would the cost of that route be?

- Comment on the optimality of the two A* algorithms. How do you account for the different routes returned?

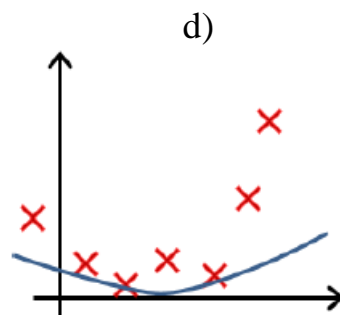
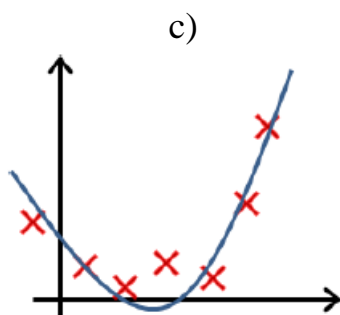
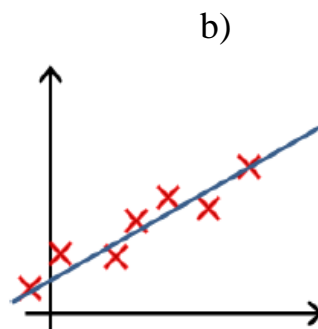
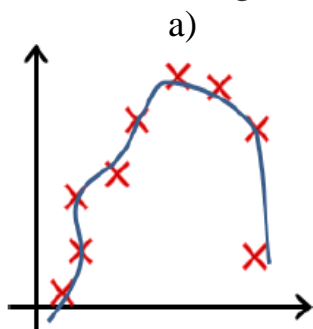
Question No. 3

(20 marks)

For each of the following, please circle the letter introducing the best answer. (Check all that apply.) Explain your answer.

- A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. In this setting, what is T?
 - The process of the algorithm examining a large amount of historical weather data.
 - The weather prediction task.
 - The probability of it correctly predicting a future date's weather.
 - None of these.

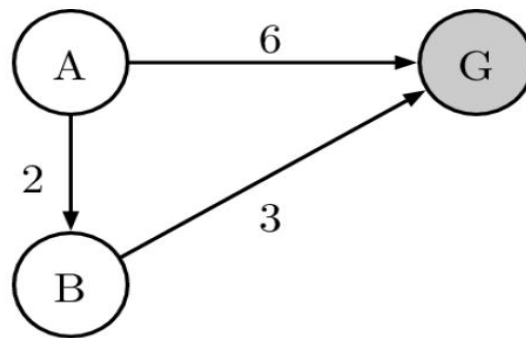
2. Suppose you are working on stock market prediction, and you would like to predict whether or not a particular stock's price will be higher tomorrow than it is today. You want to use a learning algorithm. Which one of the following algorithms is appropriate?
 - a) Regression.
 - b) Classification.
 - c) Clustering.
 - d) Reinforcement learning.
3. Suppose you have a dataset with $m=50$ examples and $n=200000$ features for each example. You want to use multivariate linear regression to fit the parameters to our data. Should you prefer gradient descent or the normal equation?
 - a) The normal equation, since gradient descent might be unable to find the optimal θ .
 - b) The normal equation, since it provides an efficient way to directly find the solution.
 - c) Gradient descent, since it will always converge to the optimal θ .
 - d) Gradient descent, since $(X^T X)^{-1}$ will be very slow to compute in the normal equation.
4. K-means is an iterative algorithm, and two of the following steps are repeatedly carried out in its inner-loop. Which two?
 - a) The cluster assignment step, where the parameters $C^{(i)}$ are updated.
 - b) Move the cluster centroids, where the centroids μ_k are updated.
 - c) Feature scaling, to ensure each feature is on a comparable scale to the others.
 - d) Using the elbow method to choose K.
5. In which one of the following figures do you think the hypothesis has overfit the training set?



6. Consider a CSP with three variables: A, B, and C. Each of the three variables can take on one of two values: either 1 or 2. There are three constraints: $A \neq B$, $C \neq B$, and $A \neq C$. Check all the choices below that are eliminated by enforcing arc consistency.
 - a) A: 1
 - b) A: 2
 - c) B: 1
 - d) B: 2
 - e) C: 1
 - f) C: 2
 - g) None of the above
7. For which of the following tasks might K-means clustering be a suitable algorithm? Select all that apply.
 - a) Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.
 - b) Given many emails, you want to determine if they are Spam or Non-Spam emails.
 - c) Given sales data from a large number of products in a supermarket, figure out which products tend to form coherent groups (say are frequently purchased together) and thus should be put on the same shelf.
 - d) Given a database of information about your users, automatically group them into different market segments.
8. Suppose that you have trained a logistic regression classifier, and it outputs on a new example a prediction $h_{\theta}(x) = 0.7$. This means (check all that apply):
 - a) Our estimate for $P(y=1/x; \theta)$ is 0.3.
 - b) Our estimate for $P(y=1/x; \theta)$ is 0.7.
 - c) Our estimate for $P(y=0/x; \theta)$ is 0.3.
 - d) Our estimate for $P(y=0/x; \theta)$ is 0.7.
9. Regarding arc consistency, For each pair of nodes, any consistent assignment to one can be extended to the other is considered as:
 - a) Node consistency.
 - b) 1-consistency.
 - c) 2-consistency.
 - d) K-consistency.
10. A navigation system that first considers all possible routes to the destination, and then selects the shortest route is described as:
 - a) Reflex agent.
 - b) Planning agent.
 - c) Co-Agent
 - d) Substituted Agent

Question No. 4**(20 marks)**

1. Prove that A* tree search with admissible heuristic is optimal.
2. For the following questions, consider the search problem shown in the figure below. It has only three states, and three directed edges. A is the start node and G is the goal node. In the table below, four different heuristic functions are defined, numbered I through IV.



	$h(A)$	$h(B)$	$h(G)$
I	4	1	0
II	5	4	0
III	4	3	0
IV	5	2	0

- a) For each heuristic function below, check if it is an *admissible* heuristic.
 - b) For each heuristic function below, check if it is a *consistent* heuristic.
 - c) Which one of the following statements about the relationship between heuristic functions III and IV is true?
 - i. Heuristic function III dominates IV.
 - ii. Heuristic function IV dominates III.
 - iii. Heuristic functions III and IV have no dominance relationship.
 - d) Which one of the following statements about the relationship between heuristic functions I and IV is true?
 - i. Heuristic function I dominates IV.
 - ii. Heuristic function IV dominates I.
 - iii. Heuristic functions I and IV have no dominance relationship.
3. Illustrate with drawing the components of an Expert Systems and Explain the Advantages & Disadvantages of them.

Question No. 5**(20 marks)**

1. Describe the basic idea of gradient descent.

2. Given the points $A = (1,2)$, $B = (2,2)$, $C = (2, 1)$, $D = (-1, 4)$, $E = (-2, -1)$ and $F = (-1,-1)$.

a) Starting from initial clusters $\text{Cluster1} = \{A\}$ which contains only the point A and $\text{Cluster2} = \{D\}$ which contains only the point D, follow the K-means clustering algorithm and report the final clusters. Use L1 distance as the distance between points which is given by

$$d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$$

b) Draw the points on a 2-D grid and check if the clusters make sense.

3. **Complete the following sentences:**

- a. is a system that employs human knowledge captured in a computer to solve problems that ordinarily require human expertise.
- b. agent choose action based on current percept and do not consider the future consequences of their actions.
- c.specify what fringe nodes do you explore next?
- d.is a function that *estimates* how close a state is to a goal.
- e. A* is optimal with and consistent heuristics.
- f. is a special subset of search problems in which state is defined by variables with values from a domain and the goal test is a set of constraints specifying allowable combinations of values for subsets of variables.
- g. Depth-first search with variable-ordering and fail-on-violation is called
- h. is keeping track of domains for unassigned variables and cross off bad options.
- i. An arc $X \rightarrow Y$ is iff for *every* x in the tail there is *some* y in the head which could be assigned without violating a constraint.
- j. is choosing the variable with the fewest legal left values in its domain.

Best wishes

Dr. Sherin El Gokhy